Modeling in the Digital Humanities: a Research Program?

Jannidis, Fotis

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Modeling in the Digital Humanities: a Research Program?

Fotis Jannidis [*]

Abstract: *»Modellierung in den Digital Humanities: ein Forschungsprogramm?«*. The term modeling is used in so many different research fields that the assumption that they are all conceptually connected seems quite hard to defend. Therefore, any research on modeling in digital humanities probably has to work inductively, to collect examples of all these different practices, in order to determine which have essential communalities. The formal modeling of research data in the humanities, up to now most often just discussed under a technical perspective, seems to be a good starting point for this endeavor.

Keywords: Data modeling, concept analysis, formal modeling.

1. Model and Modeling

It has been acknowledged that modeling is at the core of the digital humanities (McCarty 1999, Flanders and Jannidis 2015),[1] but the vagueness of the term has proven to be challenging for any attempt to build on this insight. Therefore, I will start out with investigating which concept of modeling and which of its features could be fruitful for the digital humanities. Let us start with a definition of modeling which is general enough to include some of the very different contexts the term can appear in:

> a model is a representation of something by someone for some purpose at a specific point in time. It is a representation which concentrates on some aspects - features and their relations - and disregards others. The selection of these aspects is not random but functional: it serves a specific function for an

[*] Fotis Jannidis, Institut für Deutsche Philologie, Julius-Maximilians-Universität Würzburg, Am Hubland, 97074 Würzburg, Germany; fotis.jannidis@uni-wuerzburg.de.

[1] This short essay picks up some threads, which have been started as part of a collaborative effort. Together with Julia Flanders I organized a conference on data modeling in the digital humanities which resulted in a white paper on the topic: Flanders and Jannidis 2015. We also edited a volume on data modeling which will be published in 2018 (Flanders Jannidis 2018). After so many discussions I am not able to claim any of the thoughts in this essay purely as my own.

Historical Social Research Supplement 31 (2018), 96-100 | published by GESIS
DOI: 10.12759/hsr.suppl.31.2018.96-100

> individual or a group. And a model is usually only useful and only makes sense in the context of these functions and for the time that they are needed.[2]

Models are usually expressed in some medium, they are representations. That is one of the reasons why many insights from media studies about mediality or from semiotics about signs can be applied to them. Models are not the modelled object itself, so for some aspect and under some perspective they will always distort some information, which is the basis for the famous dictum "all models are wrong".[3] But though this can be dramatized, usually this is not a problem for a successful use of a model. Comparing models to reality, one will always find some difference which can be seen as a deficiency, but this usually neglects the functional aspects of models. The importance of the relation between a model and its use cannot be overstated, because the functional requirements a model tries to fulfill are defined by this relation. Last but not least it seems useful to me to include the group which will use the model into our model of a model. This allows us to see models in a pragmatic context and distinguish between semantic aspects of a model and the details of its use.

Because this is still very general, it may help to have a closer look at what has been called a model *in research*. The term is used in natural and social sciences and the humanities to refer to the general theory, the theoretical framework behind our research, like the "standard model in particle physics" or the "theory of modernization". Or we can use it to refer to some specific part of the general model, which we want to investigate in detail, for example the Higgs boson or the establishment of a self-sustained social system like the judicial system. To be able to do research we need an operationalization, such as the experiment in the Large Hadron Collider with the collisions between protons, or the collection of historical material to show that, according to some defined indicators, there is a judicial system and it is – in some sense – self-contained. And for this part, the operationalization, the term "model" has been used too.

There is the whole field of mathematical modeling which is used at many different levels. I just mentioned the cases whereby a model is defined as a "purposeful representation of reality", usually to describe deterministic or stochastic processes of some sort.[4] These kinds of models, which are also used by those disciplines in the humanities which work with statistics, are obviously not the same as the models we see in machine learning which are used to classify or cluster data. There is, however, a relation between the two kinds of models.

---

[2] See Jannidis and Flanders: A Gentle Introduction to Data Modeling. In: Flanders and Jannidis 2018. This definition is based on Stachowiak 1973.

[3] Cf. Box 1976, 791-799.

[4] Mooney and Swift 1999, 1.

And there is another very common use of "modeling" in the humanities: when we classify an item, thus ascribing it to some class, we often refer to the classes as models. For example, if we discuss which period a literary text should be ascribed to, it is generally assumed that the periods in question are some kind of model. In the context of computer science and DH this model is sometimes called an ontology.

In the field of research, the term modeling generally refers to many different activities, so different indeed that the question has to be asked: Can we learn anything of interest, when we ask the question "what is modeling?" Or is it more productive to discuss specific activities of modeling like theory building, hypothesis creation and testing, operationalization, mathematical modeling, machine learning models, classification and ontology creation etc.? All these activities, it could be argued, are connected by a family resemblance and focusing on this may shed some light on the term modeling. It is part of the definition of Wittgenstein's concept, that the first and the last member of the family don't have to share a common trait, they have nothing in common (Wittgenstein 1976). So looking at them will provide no deeper insight into the family.

It seems to me that the situation with the concept "modeling" is very similar, maybe even worse: the term presupposes a unity, some connecting band between its different uses, but in contrast to the famous example by Wittgenstein, the term "game", "model" is not used for a group of different activities in ordinary language, but for different activities in different disciplinary terminologies: mathematical language, language of computer science, language of machine learning, language of philosophy etc. Trying to understand what modeling is, we can collect all these different uses in specific terminologies and thus create the impression of some unity, some shared sense of the word, but probably it is a questionable endeavor.

In that light we probably should change the question and choose a new line of inquiry. If modeling is a cover term which refers to very different activities in different specialized contexts, what would be the best way to learn more about them and how can their role in the digital humanities be understood? If all these fields of modeling are important for the digital humanities, which of them are especially important?

For quite some time, certainly more than three decades, the answer was: *data modeling* is especially important. Developing and applying data models for cultural heritage has been and still is a core activity in the DH community. Digital representations of cultural heritage objects were assumed to have a much longer lifespan than many other digital objects and the modeling was either done in the context of institutions which are used to think in very long time spans or it was at least influenced by this context. So, data modeling was not only a technical problem to solve, but also an institutional problem: how to keep a data model alive? And a social problem: how to build a group of people, a community of users, around it?

There seems to be only one element which distinguishes modeling in the humanities from modeling in the *digital* humanities:[5] the need to formally model the data and at least some aspect of the phenomenon which is researched. So the role of formal modeling in modeling in the humanities in general and the relation of the formal model to the phenomena in the humanities are specific to the digital humanities.

Data modeling is a kind of *formal* modeling of entities and their features and relations. For institutional reasons it is still a relatively under-researched field in computer science because there the term "data modeling" is mainly understood to be the creation of a schema for a relational database. In my opinion data modeling is a cover term for a whole range of different "meta models" like relational databases, XML, graph theory, formal grammars etc. and we have no clear understanding of how these "meta models" can be described using a unified approach based on set theory and logic. To elucidate these connections and develop this unified approach seems to me to be an important step in building a theoretical foundation for what is nowadays very often approached as a practical engineering problem.

Because there is no general theory of the (digital) humanities, there is no way to develop a theory of modeling in the digital humanities using a deductive approach. So at the moment at least, we have to work inductively and collect insights into the features and problems of modeling in this context and maybe, at some point, we will be able to identify some general patterns. Working from top down, from the abstraction to the concrete ways of modeling, expecting to see some common traits will probably not work for the reasons given above. On the contrary, we have to assume that there are no common elements. But looking at them and trying to understand their specific challenges in the digital humanities seems to me to be an interesting research program. At the very end, there will be no grand theory of modeling, but we will have a deeper insight in, for example, the challenges historical artifacts, intentional objects, pose to all these different ways and levels of modeling. We will understand how the vagueness of historical knowledge is processed in these contexts, or how the alterity of historical objects is handled.

References

Box, George. 1976. Science and Statistics. *Journal of the American Statistical Association* 71 (356): 791-9.

[5] Maybe it is worthwhile pointing out that I am using the term "digital humanities" as an hyponym to "humanities", so any DH research is a case of research in the humanities. This has the maybe surprising effect that even some very technical development of methods in the DH is research in the humanities.

Flanders, Julia, and Fotis Jannidis. 2015. Knowledge Organization and Data Modeling in the Humanities. A whitepaper. pid: urn:nbn:de:bvb:20-opus-111270. <https://www.wwp.northeastern.edu/outreach/conference/kodm2012/flanders_jannidis_datamodeling.pdf> (Accessed 1 May, 2018).
Flanders, Julia, and Fotis Jannidis, eds. 2018. *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources*. London and New York: Routledge.
McCarty, Willard. 2005. *Humanities Computing*. London and New York: Palgrave.Mooney, Douglas, and Randall Swift. 1999. *A Course in Mathematical Modeling*. The Mathematical Association of America.
Stachowiak, Herbert. 1973. *Allgemeine Modelltheorie*. Wien, New York: Springer.
Wittgenstein, Ludwig. 1976. *Philosophical Investigations*. Oxford: Blackwell.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Supplement 31 (2018):
Models and Modelling between Digital Humanities – A Multidisciplinary Perspective